

A Digital Humanities Approach to the History of Science

Eugenics revisited in hidden debates by means of semantic text mining

Pim Huijnen¹, Fons Laan² and Maarten de Rijke², and Toine Pieters¹

¹ Descartes Centre for the History and Philosophy of the Sciences and the Arts,
Utrecht University, The Netherlands

{p.huijnen, t.pieters}@uu.nl,

² ISLA, University of Amsterdam, The Netherlands

{a.c.laan, derijke}@uva.nl

Abstract. Comparative historical research on the the intensity, diversity and fluidity of public discourses has been severely hampered by the extraordinary task of manually gathering and processing large sets of opinionated data in news media in different countries. At most 50,000 documents have been systematically studied in a single comparative historical project in the subject area of heredity and eugenics. Digital techniques, like the text mining tools WAHSP and BILAND we have developed in two successive demonstrator projects, are able to perform advanced forms of multi-lingual text-mining in much larger data sets of newspapers. We describe the development and use of WAHSP and BILAND to support historical discourse analysis in large digitized news media corpora. Furthermore, we argue how text mining techniques overcome the problem of traditional historical research that only documents explicitly referring to eugenics issues and debates can be incorporated. Our tools are able to provide information on ideas and notions about heredity, genetics and eugenics that circulate in discourses that are not directly related to eugenics (e.g., sport, education and economics).

1 Introduction

The mass digitization of books, newspapers and other historical material has achieved new heights in recent years and bears the promise of exciting new possibilities for historical research. Historical scholars are increasingly incorporating computational tools and methods in all phases of their research. Digital tools are used in opening up, presenting, and curating textual and multi-media sources, in heuristic techniques of retrieval and accumulation of digitized data, in semantic text mining or geospatial information studies, in various forms of visualization and in enhanced and multi-media publications of research results, blogs, and wikis. Digital history is a methodological approach framed by the capacities of these digital tools to make, define, query, and annotate associations and analyse long term patterns of economic, technological and cultural change in the human record of the past. Digital history can be said to touch upon all aspects and forms of historical scholarship. It is not a unified field or methodology, however. It is ‘an array of convergent practices’ that come together around digitized data and digital tools [18].³ Despite inspiring examples of excellent digital historical scholarship,

³ For an overview of recent trends and discussions in digital humanities see [1, 3, 5, 28].

like mining the google books archive using ngrams [20], finding happenings ‘that never happened’ with the use of machine-learning, techniques [10] engaging in creative visual analysis of historical geography,⁴ or studying the circulation of knowledge and learned practices by means of a virtual research environment (VRE)⁵ historians have only just begun to explore what it means doing history from the perspective of both humanities and computer sciences [6, 8, 25].

Semantic text analytics is a particularly promising form of data mining that can be applied to textual data in order to derive subject-specific information out of ‘mountains’ of textual data without having to read it all. Text analytics or text mining is an umbrella term for incorporating and implementing a wide range of tools or techniques (algorithms, methods), including data mining, machine learning, natural language processing, and artificial intelligence. Semantic text analytics focuses specifically on the historical-contextual meanings of words and phrases in a big data set [12].

The goal of text mining is to reduce the effort required of humanities researchers to obtain useful information from large digitized text data sources. Text mining tools are able to mine and process large numbers of texts reasonably quickly and point researchers to discourses, sentiments, named entities or potentially meaningful concepts. Thus these tools help to reconstruct and analyse past mentalities. As evidently shared values, mentalities normally only surface when they are contested. After all, people’s mentality consists of their values, ideals and standards and these are rarely explicit. People’s worries, disputes, or what excites them are all expressions that provide an understanding of past mentalities.

Current research programs such as Digging into Data and CATCH-plus demonstrate the feasibility of performing interdisciplinary humanities research facilitated by digital research tools.⁶ Adapting digital methodologies arising from these programs to humanities research gives rise to more easily reproducible results, more refined computationally-based research methods for historians, and new research questions. These programs also demonstrate that collaborative, trans-disciplinary and integrative strategies such as common group learning (all knowledge is necessarily pooled and learning is both shared and cumulative), modeling, negotiation among experts, and integration by leaders are central to the functioning and therefore the success of this approach. The design and execution of such digital humanities programs is obviously grafted on common practice in the sciences and may be contrasted to the large majority of humanities research (exceptions excluded, notably in linguistics), where research is individualistic to the core. The role of humanities experts in the field, in our case historians of science, in the development of text and data mining technology is particularly important. This applies also to articulating and aligning the needs of users with the technological options. Incorporating regular feedback loops, for instance, allows an iterative refinement of analysis algorithms and the development of a user-friendly digital tool.

Our thesis, consequently, is that digital tools can help historians of science gain a better understanding of scientific dissemination and cultural transfer. This is what we argue in this article, as we focus on a case study of eugenics. After all, based on a wide array of undisputed scientific theories, eugenics has important cultural and politi-

⁴ <http://www.stanford.edu/group/spatialhistory> (accessed 08-09-2013)

⁵ <http://ckcc.huygens.knaw.nl/> (accessed 08-09-2013)

⁶ <http://www.diggingintodata.org>, <http://www.catchplus.nl>.

cal connotations around the world. In the words of Levine and Bashford, eugenics was not simply a sideline to our cultural heritage, but rather a central component of European modernity [9, 16, 17, 23, 27]. The highly problematic present connotation of the concept is an obvious heritage of the Second World War, and the racial laws under Nazi rule in particular. However, the sole focus on national-socialist race hygiene hardly does justice to the broad spectrum of meanings that eugenics has generated throughout history. Digital techniques can help to provide a better insight into the cultural meanings of eugenics.

2 Towards historical sentiment mining in public media

2.1 WAHSP The first step towards the development of an open source mining technology that can be used by historians without specific computer skills, is to obtain hands-on experience with research groups that use currently available open source mining tools. A recently developed tool that has been utilized to accomplish this is the CLARIN-supported web application for historical sentiment mining (a form of semantic text analytics that focuses on historical opinions, attitudes, and value judgments) in public media that is known under its acronym WAHSP.⁷ WAHSP is specifically designed for text mining the digital newspaper archive of the National Library of the Netherlands. At present, this repository comprises over 10 million pages from over 200 newspapers and periodicals published between 1618 and 1995, adding up to over 100 million articles.⁸

The technical basis of WAHSP is an ElasticSearch instance combined with the xTAS text analysis service.⁹ xTAS includes modules for online and offline processing. xTAS provides other essential text pre-processing modules (morphological normalization, format and encoding reconciliation, named entity recognition and normalization [19], etc.). It also incorporates algorithms and tools for identification of polarity (positive/support or negative/criticism), sources (opinion-holders), frequency of items and specific targets of discourses [13]. WAHSP comes with visualization modules built in D3.js (interactive word clouds and timelines). A statistical machine translation service is also available, which can be used to translate existing lexicons and documents between Dutch and German (both directions). The functionalities of xTAS are used to leverage interactive creation, expansion and refinement of lexicons specific to the user's research questions and needs. xTAS feeds visualizations that allow users to examine the research domain along the aforementioned dimensions of time, context, and the identity and frequency of the discourse. WAHSP has been developed in a specific research context, but is generic and usable in other domains in which analysis of topics, contexts and attitudes in large volumes of text is needed.

The main added value of the WAHSP tool lies in its possibilities for exploratory reading of historical patterns in public debates. The WAHSP research team found that, in terms of methodology, semi-automatic document selection fits rather well with historical research as an alternative to manual or random sampling, facilitating the combination of qualitative and quantitative approaches. Through text mining and visualization,

⁷ <http://www.clarin.nl/page/about/2>; <http://biland.nl>

⁸ The Dutch Royal Library, <http://kranten.kb.nl>.

⁹ <http://xtas.net>

new insights can be gained from an initial selection. Word clouds depicting the linguistic context within which keywords occur are instrumental in helping the historian with expert knowledge of the domain to combine and compare different historical periods in a free associative manner on the basis of a large number of historical documents. Each query immediately yields a document selection, without laborious sampling. This speeds up the heuristic process considerably. Exploring word associations and metadata, as well as visualizations of the documents over time, can lead to improved queries and, therefore, to a more representative document selection. Such quantitative analysis enhances the knowledge of the historian. A clear benefit of using exploratory searches is to allow the historian to recycle previous insights to investigate new research questions. Comparing document selections using quantitative analysis helps to validate these selections, making them less arbitrary and thus more representative.

2.2 Exploring the meaning of Eugenics in hidden debates Point of departure was the multiple discourses that converged around the use and adaptation of genetic knowledge and eugenics in the workplace, the home, and the wider world. Whereas eugenics has been the object of thorough historical research, both as an object of scientific discourse and its implementation in state policies, this is not true for the dissemination of eugenic thought in society [26]. Notions of eugenics have formed—either latently or explicitly—part of modern Western culture since the rise of genetics and hereditary science. Marius Turda has defined eugenics as ‘a cluster of social, biological and cultural ideas, centered on the redefinition of the individual and the national community according to the laws of natural selection and heredity’ [27, p. 124] In this sense, eugenics can be seen as the (pseudo-)scientific justification of possibly traditional customs and rules within communities to keep the population ‘pure’ and alive. For example, rules that prohibit marriages outside a specific community have been quite common throughout history. In the sense that theories of eugenics give an explanation and a justification for these customs, one could say that eugenics is about discourse. Moreover, as eugenic notions presuppose power relations and promote inequality we understand discourse here in its Foucauldian sense.

The challenge for historians is to both qualify and quantify the impact of genetics and eugenics on culture in their historical settings and to make these power relations explicit. The ambiguities in the meaning of heredity, as mentioned above, make this type of research even more difficult. The same can be said for the fact that what is at stake here are mentalities rather than philosophies. Scientifically defined and elaborated concepts are not of primary interest for historians in this sort of cultural history (or ‘history of mentalities’), but rather the unconscious, latent use of genetic or eugenic ideas by ordinary people [24].¹⁰ For example, in the latter part of the 19th and early part of the 20th century, in a number of Western countries, there was gradually mounting public support for “hard-line” policies of eugenics, such as marriage restrictions and involuntary sterilization. In the Netherlands, however, the official eugenics movement has always been fairly marginal; thus, “hard-line” eugenics supporters were not in the position to monopolize public discourse. What did this mean for the dissemination of

¹⁰ Following Müller-Wille en Rheinberger, genetics in this sense is ‘the transmission of physical properties in one generation to the other’: *The Production of an Epistemic Space*, 11. The minimal definition is, to quote the so-titled book by Diane Paul, ‘Controlling human heredity.’

genetic and eugenic thinking in Dutch public discourse? What were the political and racial connotations of the arguments that circulated in newspapers and how did they become manifest in various domains—for instance: negative and positive forms of inbreeding, the growing medical interest in sports, the circus or gendered questions like what defines good motherhood? It is difficult to grasp these problems by “traditional” historical means. Digital humanities methods can be of key to help answer questions like these.

2.3 Text mining within the context of tacit knowledge The WAHSP tool enables historians to collect and process large sets of opinion-text data from news media and extract discourse identity and intensity patterns—and to compare these patterns between different countries. The advantages above ‘traditional’ ways of doing historical research are clear. First, the tool enables research with no limitation on source materials. It overcomes selection issues as well as representation issues. Second, the tool makes research reproducible. Normally, an intensive research method makes it almost impossible to change or add research questions or viewpoints after analysis. One simply has to do all the work once again to ask a new question. WAHSP is made to repeat the heuristic process over and over again. Third, the tool enables research on hidden debates because one can search a combination of keywords that do not necessarily refer to specifically to eugenics, but rather imply eugenic thinking. Examples include: ‘ancestry’, ‘lineage’, ‘descent’, ‘stock’, ‘reproduction’, ‘regulation’, ‘selection’, ‘pure’/‘purity’, ‘progression’, ‘evolution’, ‘deterioration’, ‘depravation’, ‘isolation’, and ‘segregation’. Keyword searching is in itself not unproblematic—a rather “blunt” instrument in the words of Adrian Bingham [4, p. 229]. Finding the right keywords demands expert knowledge of the field of study and a great deal of perseverance and creativity [21, p. 67].

By combining these words with keywords from social or cultural domains like sports, circus, religion, and the like, one can obtain explicit discussions not only about eugenics, but also about mentalities and implicit notions influenced by hereditary and eugenic thinking within certain debates. After all, as a research object within the history of mentalities, we are not primarily interested in explicit racial and political debates influenced by genetics or eugenics. The aim of this study is, after all, to demonstrate that notions of heredity and eugenics have had an impact, no matter how diffuse or unconscious, on Dutch and German societies. Therefore, the point of departure of this research is not necessarily to identify the domains in which these debates are to be found. On the contrary, it is the identification of these domains that is the goal of this study. An essential difference of using the WAHSP application compared to traditional scholarly methods is the tool’s ability to differentiate the impact of genetics and eugenics in time and place. The tool enables the historian to make visible the relationships between “universal” scientific knowledge and its transfer within local contexts.

Like the “traditional” historian’s craftsmanship, using the WAHSP tool starts with a comprehensive overview of the field of study. Detailed knowledge of contemporary science and fashionable cultural concerns within distinct historical settings enables the researcher to formulate a number of research questions. These are turned into queries that consist of carefully selected keywords—like the examples mentioned above. In addition to keywords, a query can include a specific time period, a particular document source, or any combination of these. Working with a limitless number of queries (in theory) removes the limitation of having a single sampling strategy.

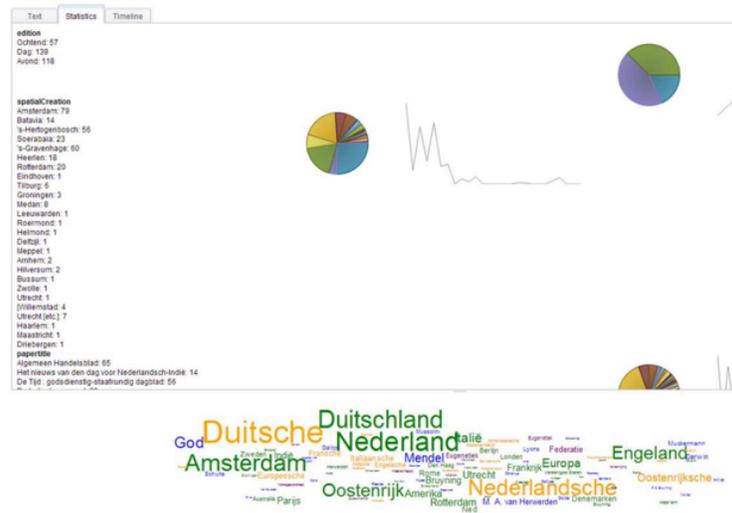


Fig. 2: WAHSP search result with named entity recognition cloud based on KB newspaper repository using query ‘eugenetiek’ (‘eugenics’) for 1900–1940.

3 Case studies

3.1 Exploratory search Self-acclaimed eugenicists were the strongest believers and successful advocates of eugenics laws and practices, as the history of many countries has shown. One reason why “hard-line” eugenics-supporters were not in the position to monopolize public discourse in the Netherlands was the absence of a strong eugenics movement. There was a strong social opposition against measures based upon eugenic notions like segregation, forced castration or sterilization. However, this opposition had more to do with a principal aversion against state intervention in what were considered private affairs than with moral problems with eugenics as such. Therefore, it remains interesting to know if the absence of eugenic practices means that eugenic thinking was absent altogether in the Netherlands.

The WAHSP tool is able to localize distinct cultural domains and public discourses that did not openly flirt with hard-line genetics, but yet had definite notions of genetic and eugenic thinking. The hints of eugenic notions in pre-war Dutch economic debates can serve as an example—one that, moreover, has turned out to be highly suitable to illustrate what we mean by exploratory search methods. The Princeton based economic historian Thomas C. Leonard argues how Progressive Era (ca. 1890-1920s) economists have advocated the minimum wage in the US as a eugenic tool. It would cause job losses and, as a consequence, it would discourage prospective immigrants to the US, as well as remove from employment the more unfit (the so-called “low wage races”): “The minimum wage protects deserving workers from the competition of the unfit by making it illegal to work for less” [14, 15, p. 213]. It is an interesting question whether similar arguments were used in Dutch debates on minimum wage. The Netherlands did not adopt a general minimum wage before 1968. Nonetheless, the introduction of a minimum wage was debated from as early as the turn of the 20th century. The WAHSP tool generates almost 10,000 hits on ‘minimum wage’ before 1945. These, obviously,

historical narratives are commonly found in reports, correspondence or manifests—i.e., in sources in which motives are generally explicated. Relying on these sources always means taking into account justifications or propaganda. This is less the case in text mining. The newspapers corpora in which the WAHSP tool searches are, obviously, by no means unbiased. However, generating results from large numbers of texts tends to lessen the effects of particular opinions. Moreover, the WAHSP tool is able to extract implicit notions and assumptions from its sources. By doing so, it bypasses biases in these sources altogether. For example, before WWII the Dutch ‘special education’ domain was surrounded by a discourse that clearly evokes associations with eugenic or at least discriminatory thinking. For children in the special education system—i.e., ‘weak’, ‘sick’ or ‘poor’ children—‘colonies’ were built where they could retreat for a period of time to gain strength [2]. Although commonly refuted by historians, it remains an interesting question whether the isolation and segregation of certain hereditary traits that were considered negative were central to the establishment of these private facilities. However, in this case WAHSP confirms the master narrative in every way. In no way do queries from the ‘special education’ domain yield results that can be linked to hereditary thinking. This supports the thesis that ‘poor relief’ motives for facilities for ‘unfit’ children were no mere justifications, but the result of sincere intentions.

The use of digital tools to confirm existing historical knowledge in this way may seem superfluous. However, it is our opinion that it is not the sole purpose of scholarly research to prove existing ideas false. To confirm present knowledge in new, quantitative, ways certainly has a value. Moreover, the use of digital tools in this sense has a significance of its own. It may help to get the use of these tools accepted as an additional means for historical research by the scholarly community. Besides, confirmatory results advantage the further development of these tool themselves, as Frederick W. Gibbs and Trevor J. Owens have stated: “As humanists continue to prove that data manipulation and machine learning can confirm existing knowledge, such techniques come closer to telling us something we don’t already know” [7].

One way of generating new knowledge is the ability of text mining to study the history of ideas and concepts (*Begriffsgeschichte*) in new ways. WAHSP is able to quantify and contextualize the historical uses of concepts. This involves a far more straightforward use of the text mining tool than its more intuitive heuristic use. E.g., mining the concept of ‘inheritance’, WAHSP can yield a timeline that indicates the frequency of the word in newspaper articles. Given the ambiguous meanings of the word ‘inheritance’—referring, e.g., not only to heredity, but also to legal and cultural forms of heritage—the tool can indicate the dominant contexts in which the concept appeared from year to year. In this instance, the WAHSP tool clearly demonstrates that the biological meaning of inheritance was dominant throughout the end of the 19th and the first half of the 20th century. However, the context in which the concept was debated did change considerably over time. E.g., the word cloud makes it plainly visible that articles containing the word ‘inheritance’, in 1867, predominantly focused on medical subjects (Fig. 4). In 1935, however, the medical context of inheritance has almost completely been replaced by a legal and racial context (Fig. 5).

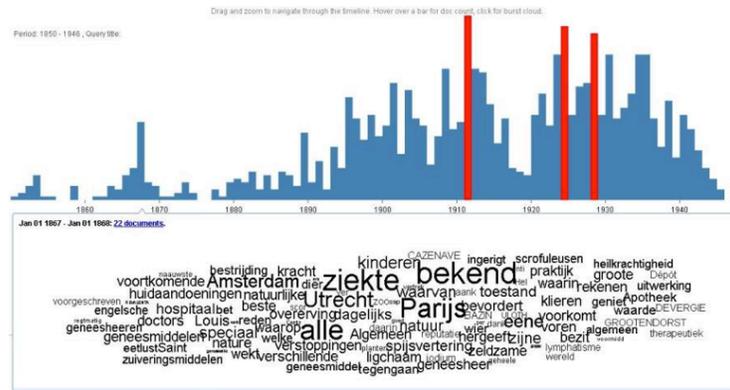


Fig. 4: WAHSP search result in the form of a timeline from the query ‘inheritance’ (‘erfelijkheid’) for 1860–1945 and an additional word cloud for the year 1867. The large number of words pointing at the medical sphere (‘doctors’, ‘hospitaal’, ‘genesmiddelen’, ‘genesheeren’, ‘apotheek’, ‘heilkrachtigheid’ etc.) gives a hint of the context with which this concept predominantly was connected.

In its relative straightforwardness, the quantification and contextualization that text mining provides offer an innovative contribution to the history of ideas and concepts.¹¹ Timelines like the ones created by the WAHSP tool show when concepts came into use, point at frequency increases and decreases and at a concept’s final disappearance. Also, the changing word clouds over time indicate the shifting contexts in which concepts were debated, thus pointing towards the varying meanings of the words in question.

3.3 New Horizons: historical text mining for comparative research As a follow-up of the WAHSP-project the bilingual text mining tool BILAND is currently being developed as an open-source and accessible web application. An interdisciplinary team of researchers are tailoring WAHSP to the language-specific needs of comparative historical research, with a particular focus on the identity, intensity, and location of discourses about heredity, genetics, and eugenics in Dutch and German newspapers between 1863 and 1940. The challenge is to incorporate the semantics of two different languages (in this case Dutch and German). As in WAHSP, BILAND employs a user-oriented, iterative model of collaboration between humanities scholars and ICT developers.

Comparative, bilingual historical text mining evokes a range of challenges. An important question is the comparability of the research topic as it is formulated in a specific query. The national vocabularies may not be literally translatable, as is, for example, the case for ‘eugenics.’ Whereas the Dutch terminology follows the English—‘eugenetica,’ ‘eugeniek’—in the German language the most common translation for ‘eugenics’ is ‘Rassenhygiene’ (‘racial hygiene’). The more literal translation ‘Eugenik’ did exist, and was used in the same sense, but would by itself in no way be a sufficient keyword to look for eugenic thought in Germany. In this specific kind of historical text mining it is, therefore, all the more important to be aware of what it is that is compared: a word (as shown in Fig. 6) or a concept, i.e., the meaning behind that word.

¹¹ As Peter Haber argues in his overview of the Digital Humanities on Docupedia-Zeitgeschichte: https://docupedia.de/zg/Digital_Humanities (24-01-2013).

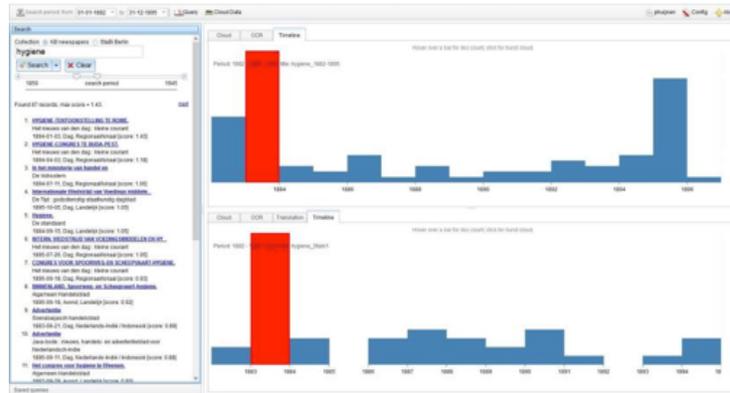


Fig. 6: BILAND search result in the form of two timelines from the query ‘hygiene’ for 1860-1900. The Dutch timeline is shown on top, the German below.

4 Conclusion

In a recent blog post called ‘The Deceptions of Data’, Andrew Prescott has criticized the jubilation of the ‘digital revolution’. He states that “One of the problems confronting data enthusiasts in the humanities is that we feel a need to convince our more old-fashioned colleagues about what can be done. But our role as advocates of digitized data shouldn’t mean that we lose our critical sense as scholars. [...] [T]here is a risk that we look more carefully at the technical components of the datasets than the historical context of the information that they represent.”¹⁶ The quintessence of Prescott’s warning is not to expect computer-generated conclusions of digital history and no machine-learned substitution for historical craftsmanship.

However, this does not mean digital techniques have little value for historians. Digital tools enable historians to analyze massive volumes of texts and other big data sets and to integrate (socio-) linguistics, statistics and geo-informatics into historical research. New techniques of large-scale data analysis allow historians to manage data sets that were only accessible by means of manual sampling. Exploratory search methods that are able to provide a quick overview combined with tools to zoom into details are especially empowering. Our proposed combination of interactive exploratory search and text mining supports historians to set up systematic search trails; the tooling helps them interpret and contrast the returned result sets: by exploring word associations for a result set, inspecting the temporal distribution of documents and by comparing selections historians can make a more informed and principled document selection. Obviously, this is no substitute for the historical workmanship. WAHSP and BILAND are meant as heuristic tools that ideally inspire new ideas and insights that would not have been generated through reading a small number of articles, but instead are only brought forth through the analysis of hundreds of articles. These insights may help to frame new research questions, thus catalyzing historical research [7]. They may help to frame new research questions, but also to stimulate serendipity. After all, digitally produced results often lead to unexpected associations that turn out promising for further research.

¹⁶ <http://digitalriffs.blogspot.nl/2013/01/the-deceptions-of-data.html>.

However, there are a number of prerequisites for the use of digital tools becoming standard procedure in historical research. First, it is quintessential that historians working with digital tools and building their arguments on digital results are highly aware of what they are doing. This may sound obvious, but it is hardly always the case. Historians should have a clear understanding of, for example, what word clouds are standing for. Or of how to translate complex queries into normal, everyday language. They should be able to interpret and explain text mining research results in formulations such as, ‘within the given source material, in all articles containing word x and word y, word z also appears with a significant frequency’. This makes their arguments transparent. As long as digital tools are treated as black boxes, with queries going in and several sorts of visualizations mysteriously coming out, the assessment of the results remains problematic. It is therefore that Gibbs and Owens argue that “[t]he processes for working with the vast amounts of easily accessible and diverse large sets of data suggest a need for historians to formulate, articulate, and propagate ideas about how data should be approached in historical research” [7]. In parallel, a thorough understanding should be developed of the search behavior of historians, in the same vein as [11].

It is essential that the status of the results from digital tools is clearly communicated. Evidently, tools like WAHSP and BILAND offer proof for certain arguments, but do not provide explanations for them. Fore-mentioned applications can show that in the Dutch public debate at the end of the 19th century, the predominant meaning of the concept of inheritance was medical, but it does not explain why.

In sum, text mining tools like WAHSP and BILAND are not built to make writing histories abundant. They are meant to trigger historians, to draw their attention to potentially interesting cases to explore. In this sense, it is evident that text mining can form a relevant addition to the historian’s toolbox outside the eugenics cases as well. It can be used to analyze trends and patterns on a much broader scale.

Acknowledgments. This research was supported by the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement nr 288024 (LiMoSIne project), the Netherlands Organisation for Scientific Research (NWO) under project nrs 640.004.802, 727.011.005, 612.001.116, HOR-11-10, the Center for Creation, Content and Technology (CCCT), the BILAND and QuaMerdes projects funded by the CLARIN-nl program, the TROVe project funded by the CLARIAH program, the Dutch national program COMMIT, the ESF Research Network Program ELIAS, the Elite Network Shifts project funded by the Royal Dutch Academy of Sciences (KNAW), the Netherlands eScience Center under project number 027.012.105 and the Yahoo! Faculty Research and Engagement Program.

5 References

- [1] Burdick et al., A.: *Digital Humanities*. MIT Press (2012)
- [2] Bakker, N.: “Kweekplaatsen van gezondheid”. *Vakantiecolonies en de medicalisering van het kindwzijn*. *Low Countries Historical Review* 126, 29–53 ((2011)
- [3] Berry, D.M. (ed.): *Understanding Digital Humanities*. Palgrave Macmillan (2012)
- [4] Bingham, A.: The digitization of newspaper archives: Opportunities and challenges for historians. *Twentieth Century British History* 21(2), 225–231 (2010)
- [5] Earheart, A.E., Jewell, A. (eds.): *The American Literature Scholar in the Digital Age*. University of Michigan Press (2011)

- [6] van Eijnatten, J., Pieters, T., Verheul, J.: Big data for global history: The transformative promise of digital humanities. *BMGN* (2014), forthcoming
- [7] Gibbs, F., Owens, T.: The hermeneutics of data and historical writing. <http://writinghistory.trincoll.edu/data/gibbs-owens-2012-spring/> (2012)
- [8] Graham, S., Milligan, I., Weingart, S.: The hermeneutics of data and historical writing. In: *The Historian's Macroscope: Big Digital History*. Imperial College Press (2013)
- [9] Hahn, D.: *Modernisierung und Biopolitik: Sterilisation und Schwangerschaftsabbruch in Deutschland nach 1945*. Campus (2000)
- [10] van den Hoven, M., van den Bosch, A., Zervanou, K.: Beyond reported history: Strikes that never happened. In: Darányi, S., Lendvai, P. (eds.) *Proceedings of the First International AMICUS Workshop on Automated Motif Discovery in Cultural Heritage and Scientific Communication Texts*, Vienna, Austria. pp. 20–28 (2010)
- [11] Huurnink, B., Hollink, L., van den Heuvel, W., de Rijke, M.: Search behavior of media professionals at an audiovisual archive: A transaction log analysis. *Journal of the American Society for Information Science and Technology* 61(6), 1180–1197 (June 2010)
- [12] Jackson, P., Moulinier, I.: *Natural Language Processing for Online Applications: Text Retrieval, Extraction and Categorization*. John Benjamins, 2nd edn. (2007)
- [13] Jijkoun, V., de Rijke, M., Weerkamp, W.: Generating focused topic-specific sentiment lexicons. In: *ACL '10* (2010)
- [14] Klausen, S., Bashford, A.: Fertility control: Eugenics, neo-malthusianism, and feminism. In: Bashford, A., Levine, S. (eds.) *The Oxford Handbook of the History of Eugenics*, pp. 98–115. Oxford University Press (2010)
- [15] Leonard, T.C.: Eugenics and economics in the progressive era. *Journal of Economic Perspectives* 19, 207–224 (2005)
- [16] Levine, P., Bashford, A.: Introduction: Eugenics and the modern world. In: Levine, P., Bashford, A. (eds.) *The Oxford Handbook of the History of Eugenics*, pp. 3–24. Oxford (2010)
- [17] Lombardo, P. (ed.): *A Century of Eugenics in America: from the Indiana Experiment to the Human Genome Era*. Indiana University Press (2001)
- [18] Lunenfeld, P., Presner, T., Schnapp, J.: Digital humanities manifesto 2.0. <http://hastac.org/node/2182> (2009)
- [19] Meij, E., Bron, M., Huurnink, B., Hollink, L., de Rijke, M.: Learning semantic query suggestions. In: *ISWC '09*. Springer (2009)
- [20] Michel, J.B.: Quantitative analysis of culture using millions of digitized books. *Science* 6014, 176–183 (2010)
- [21] Nicholson, B.: The digital turn. *Media History* 19(1), 59–73 (2013)
- [22] Odijk, D., de Rooij, O., Peetz, M.H., Pieters, T., de Rijke M., Snelders, S.: Semantic document selection. historical research on collections that span multiple centuries. In: *TPDL '12*. Springer (2012)
- [23] Reulecke, J. (ed.): *Herausforderung Bevölkerung: zu Entwicklungen des modernen Denkens über die Bevölkerung vor, im und nach dem 'Dritten Reich'*. VS Verlag für Sozialwissenschaften (2007)
- [24] Rheinberger, H.J., McLaughlin, P., Müller-Wille, S.: Introduction. In: *A Cultural History of Heredity I, 17th and 18th Century*, pp. 1–5. Preprint MPIWG, Berlin (2001)
- [25] Seefeldt, D., Thomas III, W.G.: What is digital history? a look at some exemplar projects. Faculty Publications, Department of History. Paper 98. <http://digitalcommons.unl.edu/historyfacpub/9> (2009)
- [26] Snelders, S., Pieters, T.: Van degeneratie tot individuele gezondheidsopties: Het maatschappelijk gebruik van erfelijkheidsconcepten in de twintigste eeuw. *Gewina* 26(4), 203–215 (2003)
- [27] Turda, M.: *Modernism and Eugenics*. Palgrave MacMillan (2010)
- [28] Warwick, C., Terras, M.M., Nyhan, J.: *Digital Humanities in Practice*. Facet (2012)